The many-to-few lemma and multiple spines

Simon C. Harris[1] and Matthew I. Roberts[2]

[1] *Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK.* E-mail: *S.C.Harris@bath.ac.uk*

[2] *Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany.* E-mail: *mattiroberts@gmail.com*

January 4, 2012

Abstract

We develop a simple and intuitive identity for calculating expectations of weighted k-fold sums over particles in branching processes, generalising the well-known many-to-one lemma.

1 Introduction

1.1 The many-to-two lemma

Consider a branching Brownian motion: one particle starts at 0 and moves like a Brownian motion until an exponentially distributed time with mean 1. It then dies and leaves in its place two new particles, which independently follow, relative to their initial position, the same random behaviour as their parent. Let $N(t)$ be the set of particles alive at time t, and for a particle $u \in N(t)$ let $X_u(t)$ be the position of particle u. Let B_t, $t \geq 0$ be a standard Brownian motion, and $f : \mathbb{R} \to \mathbb{R}$ a measurable function. The following result is well-known (see [3]):

Lemma 1 (Simple many-to-one lemma)**.**

$$\mathbb{E}\left[\sum_{u\in N(t)} f(X_u(t))\right] = e^t \mathbb{E}[f(B_t)]. \tag{1}$$

This lemma turns questions about a system of many dependent particles into questions about a single Brownian motion. For example, let $A(x,t) = \#\{u \in N(t) : X_u(t) > x\}$, the number of particles that are above x at time t. For which x and t is $A(x,t)$ non-zero? (This question is related to solutions of the FKPP equation.) Markov's inequality and the many-to-one lemma give us an easy upper bound:

$$\mathbb{P}(A(x,t) \geq 1) \leq \mathbb{E}[A(x,t)] = \mathbb{E}\left[\sum_{u\in N(t)} \mathbb{1}_{\{X_u(t)>x\}}\right] = \frac{e^t}{\sqrt{2\pi t}} \int_x^\infty e^{-y^2/2} dy.$$

For a lower bound, one would like to use a second-moment method, applying

$$\mathbb{P}(A(x,t) \geq 1) \geq \frac{\mathbb{E}[A(x,t)]^2}{\mathbb{E}[A(x,t)^2]}.$$

To calculate $\mathbb{E}[A(x,t)^2]$ we should use a *many-to-two* lemma.

Lemma 2 (Simple many-to-two lemma)**.** *For measurable f and g,*

$$\mathbb{E}\left[\sum_{u,v\in N(t)} f(X_u(t))\, g(X_v(t))\right] = e^{2t}\mathbb{E}[e^{T\wedge t} f(B_t)\, g(B_t')] \tag{2}$$

where

$$B_t' = \begin{cases} B_t & \text{if } t < T \\ B_T + W_{t-T} & \text{if } t \geq T \end{cases}$$

with T exponentially distributed with parameter 2 and W_t, $t \geq 0$ a standard Brownian motion independent of B_t.

The main result of this article will be the many-to-few lemma, Lemma 3, which is a much more general version of Lemma 2. In fact we will be able to calculate sums over arbitrarily many particles. We also incorporate the possibility of using a change of measure for the motion of the particles which in applications often results in easier calculations.

We prove our main result, Lemma 3, by developing a theory of multiple spines. It was not at all obvious that such a theory would even exist, let alone how to construct it so that it would lead to a clean and general many-to-few result. The single spine was introduced by Chauvin and Rouault [5], developed by Lyons *et al.* [11], and has since proved very useful in many situations in branching processes.

In special cases, results similar to Lemma 3 have existed for some time in various forms[1], usually proved by arguments specific to the particular model or problem and without any underlying spine framework. Our article provides several advantages over these previous results. We state Lemma 3 for a general model, and the spine setup gives an intuitive backdrop allowing one to transfer the result to other processes. Also, to our knowledge there is no existing work that incorporates changes of measure.

There are already several applications of this theory underway. Roberts [12] uses the full power of our general many-to-two lemma to give simple proofs of large-time asymptotics on the log scale for branching Brownian motion. Aïdékon and Harris [1] use the k-particle version to compute moments in order to show that the number of particles hitting a certain level in a branching Brownian motion with killing at the origin converges in distribution in the limit approaching criticality. Döring and Roberts [7] calculate moments of numbers of particles

[1] An even simpler form of Lemma 2 was given by Sawyer [14]. Kallenberg [10] proved a version for discrete trees, which he calls a "backward tree formula". Gorostiza and Wakolbinger [8] extend Kallenberg's formula to a class of continuous-time processes. Dawson and Perkins generate what they call "extended Palm formulas" for historical processes (superprocesses enriched with information on genealogy) in [6]. For the parabolic Anderson model with Weibull upper tails, Albeverio *et al.* [2] gave a similar result by considering existence and uniqueness of solutions to a Cauchy problem. Bansaye *et al.* [4] develop quite general many-to-two lemmas for Markov branching processes, allowing particles to be born away from their parent.

in a catalytic branching model, for which the multiple spine theory gives an intuitive combinatorial derivation for a collection of constants which otherwise appear abstractly from the analysis. Ortgiese and Roberts (work in progress) also apply the k-particle version to the parabolic Anderson model to show that the large-time behaviour of the underlying branching process is rather different from that anticipated by its moments.

The article is arranged as follows. In Section 2 we give a summary of the multi-spine setup, and then state our main result in Section 3. In Section 4 we give full constructions of the measures and filtrations used in the theory, and then prove the many-to-few lemma in Section 5. Finally, in Section 6 we state a discrete-time version of the many-to-few lemma.

2 Multiple spines

We consider a branching process starting with one particle at x under a probability measure $\mathbb{P}_x$. This particle moves within a measurable space $(J, \mathcal{B})$ according to a Markov process with generator $\mathcal{C}$. When at position y, a particle branches at rate $R(y)$, dying and giving birth to a random number of new particles with distribution μ_y (which has support on $\{0, 1, 2, \ldots\}$). Each of these particles then independently repeats the stochastic behaviour of its parent from its starting point.

We denote by $N(t)$ the set of all particles alive at time t. For a particle $u \in N(t)$ we let σ_u be the time of its birth and τ_u the time of its death, and define $\sigma_u(t) = \sigma_u \wedge t$ and $\tau_u(t) = \tau_u \wedge t$. If $u \in N(t)$ then for $s \leq t$ we write $X_u(s)$ for the position of the unique ancestor of u alive at time s. If u has 0 children then we write $X_u(s) = \Delta$ for all $t \geq \tau_u$, where $\Delta \notin J$ is a graveyard state.

2.1 The k-spine measures $\mathbb{P}^k_x$ and $\mathbb{Q}^k_x$

We define new measures $\mathbb{P}^k_x$ and $\mathbb{Q}^k_x$ under which there are k distinguished lines of descent which we call spines. Under $\mathbb{P}^k_x$ particles behave as follows:

- We begin with one particle at position x which (as well as its position) carries k marks $1, 2, \ldots, k$.
- All particles move as Markov processes with generator $\mathcal{C}$, independently of each other given their birth times and positions, just as under $\mathbb{P}_x$.
- We think of each of the marks $1, \ldots, k$ as distinguishing a particular line of descent or "spine", and define ξ^i_t to be the position of whichever particle carries mark i at time t.
- A particle at position y carrying j marks $b_1 < b_2 < \ldots < b_j$ at time t branches at rate $R(y)$, dying and being replaced by a random number of particles with law μ_y independently of the rest of the system, just as under $\mathbb{P}_x$.
- Given that a particles $v_1, \ldots, v_a$ are born at a branching event as above, the j spines each choose a particle to follow independently and uniformly at random from amongst the a available. Thus for each $1 \leq l \leq a$ and

$1 \le i \le j$ the probability that v_l carries mark i just after the branching event is $1/a$, independently of all other marks.

- If a particle carrying $j > 0$ marks $b_1 < b_2 < \ldots < b_j$ dies and is replaced by 0 particles, then its marks remain with it as it moves to the graveyard state Δ.

In other words, under $\mathbb{P}_x^k$ the system behaves exactly as under $\mathbb{P}_x$; the only difference is that some particles carry extra marks showing the lines of descent of k spines. We call the collection of particles that have carried at least one spine up to time t the *skeleton* at time t, and write $\mathrm{skel}(t)$; see Figure 1. Of course $\mathbb{P}_x^k$ is not defined on the same σ-algebra as $\mathbb{P}_x$. We let $\mathcal{F}_t^k$ be the filtration containing all information about the system (including the k spines) up to time t; then $\mathbb{P}_x^k$ is defined on $\mathcal{F}_\infty^k$. This will be clarified in Section 4.

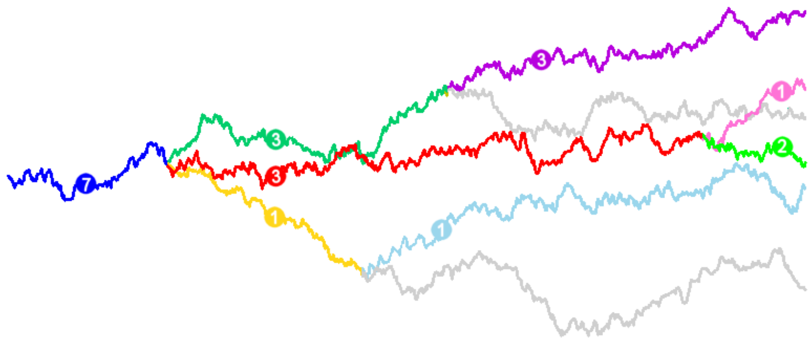

Figure 1: Each particle in the skeleton is a different colour, and particles not in the skeleton are drawn in grey. The numbers show how many spines are carried by each particle.

Now, for each $n \ge 0$ and $y \in \mathbb{R}$ let

$$m^n(y) = \sum_a a^n \mu_y(a),$$

the nth moment of the offspring distribution. Let

$$\mu_y^n(a) = \frac{a^n \mu_y(a)}{m^n(y)} \ ;$$

μ_y^n is called the nth *size-biased* distribution with respect to μ_y. For $1 \le i, j \le k$ define $T(i,j)$ to be the first split time of the ith and jth spines, i.e. the first time at which marks i and j are carried by different particles. Let $D(v)$ be the total number of marks carried by particle v.

Suppose that $\zeta(X,t)$ is a functional of a process $(X_t, t \ge 0)$ such that if $(X_t, t \ge 0)$ is a Markov process with generator $\mathcal{B}$ then $\zeta(X,t)$ is a non-negative unit-mean martingale with respect to the natural filtration of $(X_t, t \ge 0)$. For example if X is a Brownian motion on $\mathbb{R}$ then we might take $\zeta(X,t) = e^{X_t - t/2}$. Under $\mathbb{Q}_x^k$ particles behave as follows:

- We begin with one particle at position x which (as well as its position) carries k marks $1, 2, \ldots, k$.

- Just as under $\mathbb{P}_x^k$, we think of each of the marks $1, \ldots, k$ as a spine, with ξ_t^i the position of whichever particle carries mark i at time t.
- A particle with mark i at time t moves as if under the changed measure $Q_x^i|_{\sigma(\xi_s^i, s \leq t)} := \zeta(\xi^i, t)\mathbb{P}_x^k|_{\sigma(\xi_s^i, s\leq t)}$.
- A particle at position y carrying j marks at time t branches at rate $m^j(y)R(y)$, dying and being replaced by a random number of particles with law μ_y^j independently of the rest of the system.
- Given that a particles $v_1, \ldots, v_a$ are born at such a branching event, the j spines each choose a particle to follow independently and uniformly at random, just as under $\mathbb{P}_x^k$.
- Particles not in the skeleton (those carrying no marks) behave just as under $\mathbb{P}$, branching at rate $R(y)$ and giving birth to numbers of particles with law μ_y when at y.

In other words, under $\mathbb{Q}_x^k$ spine particles move as if weighted by the martingale ζ; they breed at an accelerated rate; and they give birth to size-biased numbers of children.

3 The many-to-few lemma

We note here that if Y is measurable with respect to $\mathcal{F}_t^k$, then it can be expressed as the sum

$$Y = \sum_{v_1,\ldots,v_k \in N(t)\cup\{\Delta\}} Y(v_1, \ldots, v_k)\mathbb{1}_{\{\xi_t^1 = v_1, \ldots, \xi_t^k = v_k\}}$$

where each $Y(v_1, \ldots, v_k)$ is $\mathcal{F}_t$-measurable. To see this one can generalize the argument on pages 24-25 of [13]. Since this is a purely measure-theoretic argument and will be clear for most Y of interest, we leave it as an exercise for the reader. We now state our main result in full.

Lemma 3 (Many-to-few)**.** *For any $k \geq 1$ and $\mathcal{F}_t^k$-measurable Y as above,*

$$\mathbb{P}_x\left[\sum_{v_1,\ldots,v_k\in N(t)} Y(v_1,\ldots,v_k)\mathbb{1}_{\{\zeta(v_i,t)>0\ \forall i=1,\ldots,k\}}\right]$$
$$= \mathbb{Q}_x^k\left[Y\prod_{v\in\text{skel}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\exp\left(\int_{\sigma_v(t)}^{\tau_v(t)}\left(m^{D(v)}(X_v(s))-1\right)R(X_v(s))ds\right)\right].$$

Note that under $\mathbb{Q}_x^k$, all particles $v \in \text{skel}(t)$ have $\zeta(v,t) > 0$ for all t and hence under $\mathbb{P}_x$ only those particles u with $\zeta(u,t) > 0$ can possibly be counted. This is why the indicator function appears in the $\mathbb{P}_x$-expectation on the left-hand side. In applications, if we wish to introduce a model incorporating killing of particles, this can be achieved by allowing ζ to become zero so that we only count those particles still alive at time t.

4 Multiple spines and changes of measure

Our main aim in this section is to give full details of the setup introduced in Section 2.

4.1 Trees

We use the *Ulam-Harris labelling system*: define a set of labels

$$\Omega := \{\emptyset\} \cup \bigcup_{n\in\mathbb{N}} \mathbb{N}^n.$$

We often call the elements of Ω *particles*. We think of $\emptyset$ as our inital ancestor, and a label $(3,2,7)$ for example as representing the seventh child of the second child of the third child of the initial ancestor. For a particle $u \in \Omega$ we define $|u|$, the generation of u, to be the length of u (so if $u \in \mathbb{N}^n$ then $|u| = n$, and $|\emptyset| = 0$). For two labels $u, v \in \Omega$ we write uv for the concatenation of u and v, taking $\emptyset u = u\emptyset = u$. We write $u \leq v$ and say that u is an *ancestor* of v if there exists $w \in \Omega$ such that $uw = v$.

We define $\mathbb{T}$ to be the set of all *trees*: subsets $\tau \subseteq \Omega$ such that

- $\emptyset \in \tau$: the initial ancestor is part of τ;
- for all $u, v \in \Omega$, $uv \in \tau \Rightarrow u \in \tau$: if τ contains a particle then it contains all the ancestors of that particle;
- for each $u \in \tau$, there exists $A_u \in \{0, 1, 2, \ldots\}$ such that for $j \in \mathbb{N}$, $uj \in \tau$ if and only if $1 \leq j \leq A_u$: each particle in τ has a finite number of children.

4.2 Marked trees

Since we wish to have a particular view of trees, as systems evolving in time and space, we define a *marked tree* to be a set T of triples of the form (u, l_u, X_u) such that $u \in \Omega$, the set

$$\text{tree}(T) := \{u : \exists\, l_u, X_u \text{ such that } (u, l_u, X_u) \in T\}$$

forms a tree, $l_u \in [0, \infty)$ is the *lifetime* of u, and, setting $\sigma_u := \sum_{v<u} l_v$ and $\tau_u := \sum_{v\leq u} l_u$,

$$X_u : [\sigma_u, \tau_u) \to J$$

is the *position function* of u. We think of the inital ancestor $\emptyset$ moving around in space according to its position function $X_\emptyset$ until time $l_\emptyset$. It then disappears and a number $A_\emptyset$ of new particles appear; each moves according to its position function for a period of time equal to its lifetime, before being replaced by a number of new particles; and so on.

We let $\mathcal{T}$ be the set of all marked trees, and for $T \in \mathcal{T}$ we define

$$N(t) := \{u \in \text{tree}(T) : \sigma_u \leq t < \tau_u\},$$

the set of particles alive at time t. For convenience, we extend the position path of a particle v to all times $t \in [0, \tau_v)$, to include the paths of all its ancestors:

$$X_v(t) := \begin{cases} X_v(t) & \text{if } \sigma_v \leq t < \tau_v \\ X_u(t) & \text{if } u < v \text{ and } \sigma_u \leq t < \tau_u \end{cases}$$

and if $A_v = 0$ then we write $X_v(t) = \Delta\ \forall t \geq \tau_v$.

4.3 Marked trees with spines

We now enlarge our state space further to include the notion of *spines*. A spine ξ on a marked tree τ is a subset of tree(τ) such that

- $\emptyset \in \xi$;
- $\xi \cap (N(t) \cup \{\Delta\})$ contains exactly one particle for each t;
- if $v \in \xi$ and $u < v$ then $u \in \xi$;
- if $v \in \xi$ and $A_v > 0$, then $\exists j \in \{1, \ldots, A_v\}$ such that $vj \in \xi$; otherwise $\xi \cap N(t) = \emptyset \ \forall t \geq \tau_v$.

If $v \in \xi \cap N(t)$ then we define $\xi_t := X_v(t)$, the position of the spine at time t. Sometimes we shall use the notation ξ_t to mean the particle v itself — beyond this introduction it should always be clear from the context which meaning is intended. For clarity within this section we will use the notation node(ξ_t) to denote the particle v itself — that is, the unique $v \in N(t) \cap \xi$. We say that a marked tree with spines is a sequence $(\tau, \xi^1, \xi^2, \xi^3, \ldots)$ where $\tau \in \mathcal{T}$ is a marked tree and ξ^1, ξ^2, ... are spines on τ. We let $\tilde{\mathcal{T}}$ be the set of all marked trees with spines.

4.4 Filtrations

We now work exclusively on the space $\tilde{\mathcal{T}}$ of marked trees with spines, and use different filtrations on this space to encapsulate different amounts of information. We give descriptions of these filtrations below; formal definitions are similar to those in [13] and are left to the reader.

The filtration $(\mathcal{F}_t, t \geq 0)$: We define $(\mathcal{F}_t, t \geq 0)$ to be the natural filtration of the branching process - it does not know anything about the spines.

The filtrations $(\mathcal{F}_t^k, t \geq 0)$: For each $k \geq 1$ we let $(\mathcal{F}_t^k, t \geq 0)$ be the natural filtration for the branching process and the first k spines. It does not know anything about spines ξ^{k+1}, ξ^{k+2}, ...

The filtrations $(\mathcal{G}_t^j, t \geq 0)$: For each j we define $\mathcal{G}_t^j := \sigma\left(\xi_s^j, s \in [0,t]\right)$, where ξ_s^j represents the position of the jth spine at time s. $\mathcal{G}_t^j$ contains just the spatial information about the jth spine up to time t (and whether or not it has died), but does not know which *nodes* of the tree actually make up that spine.

The filtrations $(\tilde{\mathcal{G}}_t^{\{i_1,\ldots,i_j\}}, t \geq 0)$: For each j-tuple $i_1, \ldots, i_j$ we define

$$\tilde{\mathcal{G}}_t^{\{i_1,\ldots,i_j\}} := \sigma\left(\mathcal{G}_t^k \cup \mathcal{A}_t^k \cup \mathcal{C}_t^k, k \in \{i_1, \ldots, i_j\}\right).$$

where

$$\mathcal{A}_t^k = \{\{u = \text{node}(\xi_s^k)\} : u \in \Omega, s \in [0,t]\}$$

and

$$\mathcal{C}_t^k = \{\{u < \text{node}(\xi_t^k), A_u = a, \sigma_u \leq \sigma\} : u \in \Omega, a \geq 2, \sigma \in [0,\infty)\}.$$

$\tilde{\mathcal{G}}_t^{\{i_1,\ldots,i_j\}}$ contains all the information about the relevant collection of spines up to time t: which nodes make up the spines, their positions, and for all spine nodes not in $N(t)$ (so all the strict ancestors of the spines at time t) their lifetimes and number of children.

The filtration $(\tilde{\mathcal{G}}^k_t, t \geq 0)$**:** We use the shorthand $\tilde{\mathcal{G}}^k_t = \tilde{\mathcal{G}}^{\{1,\ldots,k\}}_t$, so that $\tilde{\mathcal{G}}^k_t$ knows everything about the first k spines up to time t. Thus $\tilde{\mathcal{G}}^k_t$ is different from $\tilde{\mathcal{G}}^{\{k\}}_t$.

4.5 Probability measures

We may now take a probability measure $\mathbb{P}_x$ on $\tilde{\mathcal{T}}$ such that under $\mathbb{P}_x$, the system evolves as a branching process starting with one particle at x, each particle moves as a Markov process with generator $\mathcal{C}$ independently of all others given its birth time and position, and a particle at position y branches at rate $R(y)$ into a random number of particles with distribution μ_y. This is the system described in Section 2. This measure, however, has no knowledge of the spines (since it sees only the filtration $\mathcal{F}_t$). We would like to extend this to a measure on each of the finer filtrations $\tilde{\mathcal{F}}^k_t$. To do this, we imagine each spine, at each fission event, choosing uniformly from the available children. Then it is easy to see that, for any particle u in a marked tree T and any $j \geq 1$, we would like

$$\mathrm{Prob}(u \in \xi^j) = \prod_{v<u} \frac{1}{A_v}.$$

We recall from Section 2 that if Y is an $\tilde{\mathcal{F}}^k_t$-measurable random variable then we can write:

$$Y = \sum_{v_1,\ldots,v_k \in N(t)\cup\{\Delta\}} Y(v_1,\ldots,v_k)\mathbb{1}_{\{\xi^1_t = v_1,\ldots,\xi^k_t = v_k\}} \tag{3}$$

where each $Y(v_1,\ldots,v_k)$ is $\mathcal{F}_t$-measurable. (Here when we write ξ^j_t we are talking really about the particle node(ξ^j_t) rather than its position.)

Definition 4. We define the probability measure $\mathbb{P}^k_x$ on $(\tilde{\mathcal{T}}, \tilde{\mathcal{F}}_\infty)$, by setting

$$\mathbb{P}^k_x[Y] = \mathbb{P}_x\left[\sum_{v_1,\ldots,v_k \in N(t)\cup\{\Delta\}} Y(v_1,\ldots,v_k)\prod_{j=1}^k \prod_{u<v_j} \frac{1}{A_u}\right] \tag{4}$$

for each $\mathcal{F}^k_t$-measurable Y with representation (3). Note that $\mathbb{P}_x = \mathbb{P}^k_x|_{\mathcal{F}_\infty}$.

In summary, particles carrying spines behave just as they would under $\mathbb{P}_x$, and when such a particle branches, each spine makes an independent choice uniformly from amongst the available children.

4.6 Martingales and a change of measure

As in Section 2 define $T(i,j) := \inf\{t \geq 0 : \xi^i_t \neq \xi^j_t\}$, and suppose that we are given a functional $\zeta(\cdot, t)$, $t \geq 0$, such that $\zeta(Y,t)$ is a non-negative unit-mean martingale with respect to the natural filtration of the Markov process $(Y_t, t \geq 0)$ with generator $\mathcal{C}$. We call ζ the single-particle martingale.

Recall that we defined skel$^k(t)$, the skeleton, to be the subtree up to time t generated by those particles carrying at least one of the k spines,

$$\mathrm{skel}^k(t) = \{u \in \Omega : \exists s \leq t, j \leq k \text{ such that node}(\xi^j_s) = u\}.$$

We also set
$$D^k(v) = \#\{j \le k : \exists t \text{ with } v = \xi_t^j\}$$
to be the number of spines following particle v, and define
$$E^k(v,t) = \exp\left(- \int_{\sigma_v(t)}^{\tau_v(t)} \left(m^{D(v)}(X_v(s)) - 1 \right) R(X_v(s)) ds \right).$$
Since we will not always know which particles are the spines (when we are working on $\mathcal{F}_t$ for example), it will sometimes be helpful to have the above concepts defined for a general skeleton of k particles $u_1, \ldots, u_k$ instead of the spines. For this reason we define
$$\text{skel}_{u_1,\ldots,u_k}(t) = \{v \in \Omega : \sigma_v \le t, \exists j \text{ with } v \le u_j\},$$
$$D_{u_1,\ldots,u_k}(v) = \#\{j : v \le u_j\},$$
and
$$E_{u_1,\ldots,u_k}(v,t) = \exp\left(- \int_{\sigma_v(t)}^{\tau_v(t)} \left(m^{D_{u_1,\ldots,u_k}(v)}(X_v(s)) - 1 \right) R(X_v(s)) ds \right)$$
so that
$$\text{skel}^k(t) = \text{skel}_{\xi_t^1,\ldots,\xi_t^k}(t), \ \ D^k(v) = D_{\xi_{\sigma_v}^1,\ldots,\xi_{\sigma_v}^k}(v) \ \text{ and } \ E^k(v,t) := E_{\xi_{\sigma_v}^1,\ldots,\xi_{\sigma_v}^k}(v,t).$$
Remark. We note that, with the notation given above,
$$\mathbb{P}_x^k(\xi_t^1 = u_1, \ldots, \xi_t^k = u_k | \mathcal{F}_t) = \prod_{v \in \text{skel}_{u_1,\ldots,u_k}(t) \setminus N(t)} A_v^{-D_{u_1,\ldots,u_k}(v)}.$$

Definition 5. We define an $\tilde{\mathcal{F}}_t^k$-adapted (and, in fact, $\tilde{\mathcal{G}}_t^k$-adapted) process $\tilde{\zeta}^k(t)$, $t \ge 0$ by
$$\tilde{\zeta}^k(t) = \mathbb{1}_{\{\zeta(\xi^i,t)>0 \ \forall i=1,\ldots,k\}} \prod_{v \in \text{skel}^k(t)} \left(\frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E^k(v,t) \right) \prod_{v \in \text{skel}^k(t) \setminus N(t)} A_v^{D^k(v)}$$
(if $A_v = 0$ then we define $\zeta(X_v, \tau_v(t)) = 0$) and an $\mathcal{F}_t$-adapted process $Z^k(t)$, $t \ge 0$ by
$$Z^k(t) = \sum_{u_1,\ldots,u_k \in N(t)} \mathbb{1}_{\{\zeta(u_i,t)>0 \ \forall i=1,\ldots,k\}} \prod_{v \in \text{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t).$$

We remark here that Z^k and $\zeta(\xi^j, \cdot)$ are, in fact, simply the projections of $\tilde{\zeta}^k$ onto the relevant filtrations:
$$Z^k(t) = \mathbb{P}_x^k[\tilde{\zeta}^k(t)|\mathcal{F}_t] \quad \text{ and } \quad \zeta(\xi^j, t) = \mathbb{P}_x^k[\tilde{\zeta}^k(t)|\mathcal{G}_t^{\{j\}}].$$
Lemma 6. *The process $\tilde{\zeta}^k(t)$, $t \ge 0$ is a martingale with respect to the filtrations $\tilde{\mathcal{G}}_t^k$ and $\tilde{\mathcal{F}}_t^k$.*

Proof. Let $\chi = (v_1, v_2, \ldots)$ be a single line of descent (so in particular $v_1 < v_2 < \ldots$), with χ_t representing the position of the unique v_i that is alive at time t. The births along χ form a Cox process driven by χ_t with rate function R. Thus for any $j \geq 0$,

$$\mathbb{P}_x\left[\prod_{v<\chi_t} A_v^j \middle| \chi_s, s \in [0,t]\right] = \exp\left(\int_0^t (m^j(\chi_s) - 1)R(\chi_s)ds\right).$$

We work by induction on k. The case $k = 1$ is just the single spine case, and is proved by conditioning first on $\mathcal{G}_t^1$, since the births along the spine form a Cox process driven by ξ_t^1 with rate function R. Then, by induction, it is enough to consider the process up to the first split time of the skeleton, since after this time no particle carries more than $k-1$ spines. But up to the first split we have a single particle carrying k spines, so the same argument holds as for the single spine case: the births again form a Cox process driven by ξ_t^1 with rate function R. □

Definition 7. We define the measure $\mathbb{Q}_x^k$ by setting

$$\left.\frac{d\mathbb{Q}_x^k}{d\mathbb{P}_x^k}\right|_{\mathcal{F}_t^k} = \tilde{\zeta}^k(t).$$

The proof that $\mathbb{Q}_x^k$ behaves as claimed in Section 2.1 is identical to the proof for one spine given by Chauvin and Rouault [5], applied to each branch of the skeleton independently.

5 Proof of the many-to-few lemma

We first calculate the probability that particles $(u_1, \ldots, u_k)$ make up the skeleton at time t.

Lemma 8 (Gibbs-Boltzmann weights for $\mathbb{Q}^k$)**.** *For any $u_1, \ldots u_k \in N(t) \cup \{\Delta\}$,*

$$\mathbb{Q}_x^k(\xi_t^1 = u_1, \ldots, \xi_t^k = u_k | \mathcal{F}_t) = \frac{1}{Z(t)} \prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t).$$

Proof. By the fact that $\mathbb{P}_x^k[\tilde{\zeta}(t)|\mathcal{F}_t] = Z(t)$ and standard properties of conditional expectation,

$$\begin{aligned}
\mathbb{Q}_x^k(\xi_t^1 = u_1, \ldots, \xi_t^k = u_k | \mathcal{F}_t) &= \frac{\mathbb{P}_x^k[\tilde{\zeta}(t)\mathbb{1}_{\{\xi_t^1 = u_1,\ldots,\xi_t^k=u_k\}}|\mathcal{F}_t]}{\mathbb{P}_x^k[\tilde{\zeta}(t)|\mathcal{F}_t]} \\
&= \frac{1}{Z(t)}\left(\prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t)\right) \\
&\quad \cdot \left(\prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)\setminus N(t)} A_v^{D_{u_1,\ldots,u_k}(v)}\right) \mathbb{P}_x^k(\xi_t^1 = u_1, \ldots, \xi_t^k = u_k | \mathcal{F}_t) \\
&= \frac{1}{Z(t)} \prod_{v \in \mathrm{skel}_{u_1,\ldots,u_k}(t)} \frac{\zeta(X_v, \tau_v(t))}{\zeta(X_v, \sigma_v(t))} E_{u_1,\ldots,u_k}(v,t). \quad \square
\end{aligned}$$

The proof of the many-to-few lemma is now straightforward.

Proof of Lemma 3. We begin with the right-hand side.

$$\begin{aligned}
&\mathbb{Q}_x^k\left[Y \prod_{v\in\mathrm{skel}(t)} \frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E(v,t)}\right]\\
&= \mathbb{Q}_x^k\left[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}} Y(u_1,\dots,u_k)\prod_{v\in\mathrm{skel}_{u_1,\dots,u_k}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{1}{E_{u_1,\dots,u_k}(v,t)}\mathbb{1}_{\{\xi_t^1=u_1,\dots,\xi_t^k=u_k\}}\right]\\
&= \mathbb{Q}_x^k\left[\sum_{u_1,\dots,u_k\in N(t)\cup\{\Delta\}} Y(u_1,\dots,u_k)\prod_{v\in\mathrm{skel}_{u_1,\dots,u_k}(t)}\frac{\zeta(X_v,\sigma_v(t))}{\zeta(X_v,\tau_v(t))}\frac{\mathbb{Q}_x^k(\xi_t^1=u_1,\dots,\xi_t^k=u_k|\mathcal{F}_t)}{E_{u_1,\dots,u_k}(v,t)}\right]\\
&= \mathbb{Q}_x^k\left[\frac{1}{Z^k(t)}\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\right]\\
&= \mathbb{Q}_x^k\left[\frac{1}{Z^k(t)}\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\mathbb{1}_{\{\zeta(u_i,t)>0\ \forall i=1,\dots,k\}}\right]\\
&= \mathbb{P}_x^k\left[\sum_{u_1,\dots,u_k\in N(t)} Y(u_1,\dots,u_k)\mathbb{1}_{\{\zeta(u_i,t)>0\ \forall i=1,\dots,k\}}\right]
\end{aligned}$$

where for the last step we used the fact that $\left.\frac{d\mathbb{Q}_x^k}{d\mathbb{P}_x^k}\right|_{\mathcal{F}_t} = Z^k(t)$. □

6 Many-to-few in discrete time

We state here a version of the many-to-few lemma for discrete-time processes. We shall not prove it, as it is very similar to the continuous-time version studied above.

We begin with one particle in generation 0 located at $x \in J$. Any particle at position y has children whose number and positions are decided according to a finite point process $\mathcal{D}_y$ on J. The children of particles in generation n make up generation $n+1$. We define $N(n)$ to be the total number of particles in generation n, and X_v to be the position of particle v. We set $m^j(y) = \mathbb{P}_y[N(1)^j]$ to be the jth moment of the number of particles created by the point process $\mathcal{D}_y$. Write $|v|$ to be the generation of particle v. For a particle v in generation $n \geq 1$, let $p(v)$ be its parent in generation $n-1$. For any line of descent $v_0, v_1, v_2, \dots$ such that $|v_n| = n$ and $p(v_{n+1}) = v_n$ for each $n \geq 0$, we note that $X_{v_0}, X_{v_1}, X_{v_2}, \dots$ is a Markov chain with some generator $\mathcal{C}'$ not depending on the choice of $v_0, v_1, \dots$. Suppose that $\zeta(X,n)$, $n \geq 0$ is a functional of a process $(X_n, n \geq 0)$ such that if $(X_n, n \geq 0)$ is a Markov process with generator $\mathcal{C}'$ then $\zeta(X,n)$, $n \geq 0$ is a martingale with respect to the natural filtration of $(X_n, n \geq 0)$.

6.1 The measure $\mathbb{Q}_x^k$ and the main result in discrete time

We have k distinguished lines of descent just as in the continuous-time case, which we call spines. Under $\mathbb{P}$, if a particle carrying j marks (i.e. the particle is

part of j spines) in generation n has l children in generation $n+1$, then each of its j marks chooses a particle to follow in generation $n+1$ uniformly at random from the l children. We let ξ_n^i be the position of the ith spine in generation n and define $\mathrm{skel}(n)$ to be the set of all particles of generation at most n which are part of at least one spine. Set D_v to be the number of marks carried by particle v.

Under $\mathbb{Q}_x^k$ particles behave as follows:

- A particle at position y carrying j marks has children whose number and positions are decided by a point process such that:
  - for each j and $l \geq 0$, $\mathbb{Q}_y^j(N(1)=l) = l^j \mathbb{P}_y(N(1)=l)/\mathbb{P}_y[N(1)^j]$;
  - for each i, the sequence $X_{\xi_0^i}, X_{\xi_1^i}, X_{\xi_2^i}, \ldots$ is a Markov chain distributed as if under the changed measure $Q_x^i|_{\mathcal{G}_n^{\{i\}}} := \zeta(\xi^i, n)\mathbb{P}_x^k|_{\mathcal{G}_n^{\{i\}}}$.
- Given that a particles $v_1, \ldots, v_a$ are born at such a branching event, the j spines each choose a particle to follow independently and uniformly at random, just as under $\mathbb{P}_x^k$.
- Particles not in the skeleton (those carrying no marks) have children according to the point process $\mathcal{D}_y$ when at position y, just as under $\mathbb{P}$.

In other words, under $\mathbb{Q}_x^k$ spine particles move as if weighted by the martingale ζ, and they give birth to size-biased numbers of children.

Lemma 9 (Many-to-few in discrete time)**.** *For any $k \geq 1$ and $\mathcal{F}_n^k$-measurable Y such that*

$$Y = \sum_{v_1,\ldots,v_k \in N(n)\cup\{\Delta\}} Y(v_1,\ldots,v_k)\mathbb{1}_{\{\xi_n^1=v_1,\ldots,\xi_n^k=v_k\}}$$

we have

$$\mathbb{P}_x\left[\sum_{v_1,\ldots,v_k\in N(n)} Y(v_1,\ldots,v_k)\mathbb{1}_{\{\zeta(v_i,n)>0\ \forall i=1,\ldots,k\}}\right] = \mathbb{Q}_x^k\left[Y \prod_{v\in \mathrm{skel}(n)\setminus\{\emptyset\}} \frac{\zeta(p(v),|v|-1)}{\zeta(v,|v|)} m^{D_{p(v)}}(X_{p(v)})\right].$$

Acknowledgements

The authors are indebted to both Elie Aïdékon and Julien Berestycki for some very helpful discussions, and also thank Pascal Maillard for checking an earlier draft.

References

[1] E. Aïdékon and S.C. Harris. Near-critical survival probability of branching Brownian motion with an absorbing barrier. To appear, 2011.

[2] S. Albeverio, L.V. Bogachev, S.A. Molchanov, and E.B. Yarovaya. Annealed moment Lyapunov exponents for a branching random walk in a homogeneous random branching environment. *Markov Process. Related Fields*, 6(4):473–516, 2000.

[3] K.B. Athreya and P.E. Ney. *Branching Processes.* Springer-Verlag, New York, 1972.

[4] V. Bansaye, J. F. Delmas, L. Marsalle and V. C. Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees To appear in *Ann. Appl. Probab.* Arxiv preprint: `http://arxiv.org/abs/0911.1973`, 2010.

[5] B. Chauvin and A. Rouault. KPP equation and supercritical branching Brownian motion in the subcritical speed area. Application to spatial trees. *Probab. Theory Related Fields*, 80(2):299–314, 1988.

[6] D.A. Dawson and E.A. Perkins. *Historical processes.* AMS Bookstore, 1991.

[7] L. Döring and M. I. Roberts. Catalytic branching processes via spine techniques and renewal theory. Arxiv preprint: `http://arxiv.org/abs/1106.5428`, 2011.

[8] L.G. Gorostiza and A. Wakolbinger. Persistence criteria for a class of critical branching particle systems in continuous time. *Ann. Probab.*, 19(1):266–288, 1991.

[9] R. Hardy and S. C. Harris. A spine approach to branching diffusions with applications to L^p-convergence of martingales. In *Séminaire de Probabilités, XLII*, volume 1979 of *Lecture Notes in Math.* Springer, Berlin, 2009.

[10] O. Kallenberg. Stability of critical cluster fields. *Mathematische Nachrichten*, 77(1):7–43, 1977.

[11] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[12] M. I. Roberts. A simple path to asymptotics for the frontier of a branching Brownian motion. Arxiv preprint: `http://arxiv.org/abs/1106.4771`, 2011.

[13] M.I. Roberts. *Spine changes of measure and branching diffusions.* PhD thesis, University of Bath, 2010. Available online: `http://people.bath.ac.uk/mir20/thesis.pdf`.

[14] S. Sawyer. Branching diffusion processes in population genetics. *Advances in Appl. Probability*, 8(4):659–689, 1976.